The interspecific fungal hybrid *Verticillium longisporum* displays sub-genome-specific gene expression

Jasper R.L. Depotter[abc], Fabian van Beveren[a¶], Luis Rodriguez-Moreno[ae¶], H. Martin Kramer[a¶], Edgar A. Chavarro Carrero[a¶], Gabriel L. Fiorin[a], Grardy C.M. van den Berg[a], Thomas A. Wood[b&], Bart P.H.J. Thomma[ac&*], Michael F. Seidl[ad&*]

[a]Laboratory of Phytopathology, Wageningen University & Research, Wageningen, The Netherlands

[b]Department of Crops and Agronomy, National Institute of Agricultural Botany, Cambridge, United Kingdom

[c]University of Cologne, Institute for Plant Sciences, Cluster of Excellence on Plant Sciences (CEPLAS), Cologne, Germany

[d]Theoretical Biology & Bioinformatics, Utrecht University, Utrecht, the Netherlands

[e]Departamento de Biología Celular, Genética y Fisiología, Universidad de Málaga, Málaga, Spain

*Corresponding authors

E-mail: bthomma@uni-koeln.de (BPHJT)

E-mail: m.f.seidl@uu.nl (MFS)

[¶]These authors contributed equally to this work

[&]These authors contributed equally to this work

Abstract

Hybridization is an important evolutionary mechanism that can enable organisms to adapt to environmental challenges. It has previously been shown that the fungal allodiploid species *Verticillium longisporum*, causal agent of Verticillium stem striping in rape seed, has originated from at least three independent hybridization events between two haploid *Verticillium* species. To reveal the impact of genome duplication as a consequence of the hybridization, we studied the genome and transcriptome dynamics upon two independent *V. longisporum* hybridization events, represented by the hybrid lineages “A1/D1” and “A1/D3”. We show that the *V. longisporum* genomes are characterized by extensive chromosomal rearrangements, including between parental chromosomal sets. *V. longisporum* hybrids display signs of evolutionary dynamics that are typically associated with the aftermath of allodiploidization, such as haploidization and a more relaxed gene evolution. Expression patterns of the two sub-genomes within the two hybrid lineages are more similar than those of the shared A1 parent between the two lineages, showing that expression patterns of the parental genomes homogenized within a lineage. However, as genes that display differential parental expression *in planta* do not typically display the same pattern *in vitro*, we conclude that sub-genome-specific responses occur in both lineages. Overall, our study uncovers the genomic and transcriptomic plasticity during evolution of the filamentous fungal hybrid *V. longisporum* and illustrate its adaptive potential.

Importance

Verticillium is a genus of plant-associated fungi that include a handful of plant pathogens that collectively affect a wide range of hosts. On several occasions, haploid *Verticillium* species hybridized into the stable allodiploid species *Verticillium longisporum*, which is, in contrast to haploid *Verticillium* species, a Brassicaceae specialist. Here, we studied the evolutionary genome and transcriptome dynamics of *V. longisporum* and the impact of the hybridization. *V. longisporum* genomes display a mosaic structure due do genomic rearrangements between the parental chromosome sets. Similar to other allopolyploid hybrids, *V. longisporum* displays an ongoing loss of heterozygosity and a more relaxed gene evolution. Also, differential parental gene expression is observed, with an enrichment for genes that encode secreted proteins. Intriguingly, the majority of these genes displays sub-genome-specific responses under differential growth conditions. In conclusion, hybridization has incited the genomic and transcriptomic plasticity that enables adaptation to environmental changes in a parental allele-specific fashion.

Key words: allopolyploidization, Verticillium stem striping, genome rearrangements, gene conversion, haploidization, mosaic genome, chromatin conformation capture (Hi-C)

Introduction

Upon hybridization, two distinct genotypes are merged in a single organism. This surge in genomic variation can increase the adaptive potential of hybrid organisms, which may explain why stable hybrids are generally fitter than their parents in particular environments (1). However, hybrids may also encounter incompatibilities between parental genomes as they lack the recently shared evolutionary history (2). Hybridization can lead to the emergence of new species that are reproductively isolated from their parents, known as hybrid speciation (3, 4). Although the incidence of hybridization may be rare due to such incompatibilities, many organisms encountered hybridization at a particular point in their evolution (5). Hybridization has also impacted the evolution of humans, as our genomes still contain traces from Neanderthal introgression (6). Hybridization can occur between gametes after a conventional meiosis, leading to so-called homoploid hybrids. Alternatively, when compete sets of parental chromosomes combine, the hybridization is accompanied by genome duplication during so-called allopolyploidization.

Hybridization has impacted the evolution of a wide diversity of fungi (7–9). For instance, the yeast *Saccharomyces paradoxus*, a close relative of the baker's yeast *Saccharomyces cerevisiae*, has naturally hybridized in North America forests (10), whereas also *S. cerevisiae* itself was shown to have undergone an ancient interspecies hybridization (11). Similarly, various *Candida* species that are opportunistic human pathogens display genomic traces of hybridization events (12–15). Hybridization also contributed to the evolution of various plant pathogenic fungi (7). Plant pathogens generally co-evolve with their hosts to evade host immunity, while hosts attempt to intercept pathogen ingress (16). In this process, plant pathogens secrete effector proteins that contribute to host immunity evasion and interfere with host metabolic processes (17), or affect to other processes to contribute to host colonization (18), such as the manipulation of host microbiomes (19, 20).

Due to the increased adaptation potential, hybridization has been proposed as a potent driver in pathogen evolution as it can impact host interactions through increased virulence and host range alterations (8). For instance, the Ug99 strain of the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* arose from a hybridization event and caused devastating epidemics in Africa and the Middle East (21, 22). Recent hybridization between the wheat powdery mildew, *B. graminis* f. sp. *tritici*, and rye powdery mildew, *B. graminis* f. sp. *secalis*, gave rise to the novel mildew species *Blumeria graminis* f. sp. *triticale* that, in contrast to its parents, is able to cause disease on triticale (23).

Upon hybridization, genomes typically experience a so-called “genome shock”, inciting major genomic reorganizations that can manifest by genome rearrangements, extensive gene loss, transposon activation, and alterations in gene expression (24, 25). Conceivably, these early stage alterations are primordial for hybrid survival, as divergent evolution is principally associated with incompatibilities between the parental genomes (26). Additionally, these initial re-organizations and further alterations in the aftermath of hybridization provide a source for environmental adaptation. Frequently, hybrid genomes lose their heterozygosity over time (27). Hybrids that are still sexually compatible with one of its parents can lose heterozygosity through backcrossing. Alternatively, heterozygosity can be a result of the direct loss of a homolog of one of the two parents (i.e. a homeolog) through deletion or through gene conversion whereby one of the copies substitutes its homeologous counterpart. Gene conversion and the homogenization of complete chromosomes played a pivotal role in the evolution of the osmotolerant yeast species *Pichia sorbitophila* (28). Two of its seven chromosome pairs consist of partly heterozygous and partly homozygous sections, whereas two chromosome pairs are completely homozygous. Gene conversion may eventually result in chromosomes consisting of sections of both parental origins, so called “mosaic genomes” (29). However, mosaic genomes can also arise through recombination

between chromosomes of the different parents, such as in the hybrid yeast *Zygosaccharomyces parabailii* (30). Hybridization associated with polyploidy, allopolyploids can have an additional adaptive potential through the presence of an additional copy for most genes, which gives leeway to functional diversification (31, 32). Hybridization typically entails also alterations of gene expression patterns that are non-additive from the parental expression patterns (33, 34). Nevertheless, expression patterns are generally conserved upon hybridization, as the majority of allopolyploid genes are expressed in a similar fashion as their parental orthologs (35). For instance, more than half of the genes in an allopolyploid strain of the fungal grass endophyte *Epichloë* retained their parental gene expression pattern (36). Similar conservation has also been observed for *Blumeria graminis* f. sp. *triticale* as over half of the 5% most highly expressed genes are shared with both of its hybridization parents (37). In conclusion, the genomic and transcriptomic alterations accompanied with hybridization make that hybrids have a high potential for environmental adaptation (8).

Within the *Verticillium* genus that comprises nine haploid species, hybridization resulted in the emergence of the species *Verticillium longisporum* (38–41). *V. longisporum* is sub-divided into three lineages, each representing a separate hybridization event (39, 41). *Verticillium* species A1 is a parent of each of the three hybrids and hybridized with *Verticillium* species D1, D2 and D3, resulting in the *V. longisporum* lineages A1/D1, A1/D2, and A1/D3, respectively. Whereas species D2 and D3 have been classified as 'likely *V. dahliae*', species D1 has been classified as an enigmatic species that is closely related to *V. dahliae* (39). Species A1 is also an enigmatic species that diverged earlier from *V. dahliae* than the D1 species (39). Similar as the haploid *Verticillium* species, *V. longisporum* is thought to mainly undergo asexual reproduction, as a sexual cycle has never been described and populations are not outcrossing (40, 41). Interestingly, *V. longisporum* mainly infects plant hosts of the Brassicaceae family whereas other *Verticillium* species do not cause disease

136 on Brassicaceous hosts (42). Moreover, while *V. dahliae* is characterized by an extremely
137 broad host range that comprises hundreds of (non-Brassicaceae) plant species, *V.*
138 *longisporum* only has a limited host range and hardly infects non-Brassicaceae species (42).
139 After hybridization, *V. longisporum* conceivably encountered extensive genetic and
140 transcriptomic alterations that facilitated its viability of a hybrid and the shift towards
141 Brassicaceous hosts. In this study, we investigated the impact of allodiploidization on the
142 evolution of *V. longisporum* by investigating genome, gene, and transcriptomic plasticity
143 within and between two of the hybridization events.

Results

Verticillium longisporum displays a mosaic genome structure

The genomes of three *V. longisporum* strains from two different hybridization events were analyzed to investigate the impact of hybridization on genome structure. Previously, *V. longisporum* strains VLB2 and VL20, both belonging to the A1/D1 hybridization event, were sequenced with the PacBio RSII platform and assembled *de novo* (40). We now additionally sequenced the *V. longisporum* strain PD589, which originates from the A1/D3 hybridization event (39), using Oxford Nanopore sequencing technology and the BGISeq platform to obtain long reads and paired-end short reads, respectively. All *V. longisporum* genome assemblies were improved using chromatin conformation capture (Hi-C) sequencing that detects DNA-interactions (43). Moreover, centromeres can be located with Hi-C sequencing as they display strong interaction with centromeres in other chromosomes (44). We obtained genome assemblies of 72.7, 72.2 and 72.0 Mb consisting of 15, 15 and 16 pseudochromosomes for VLB2, VL20 and PD589, respectively (Figure 1A, Table 1). Every pseudochromosome contained a centromere, suggesting that the A1/D1 isolates have 15 chromosomes and that the A1/D3 isolate PD589 contains 16 chromosomes. However, chromosome 13 of strain PD589 displayed remarkably stronger DNA-interactions than the other chromosomes (Figure 1B, see green outline), as the median read coverage of chromosome 13 is 110x, whereas the read coverage is 58-70x for all other chromosomes (Figure S1). This finding suggests that chromosome 13 recently (partly) duplicated since the high sequence identity of the duplicated regions resulted in a collapsed assembly. Consequently, strain PD589 may therefore actually have 17 chromosomes in total.

Being able to determine the parental origin of individual genomic regions is elementary to investigate genome evolution in the aftermath of hybridization. As the D parents of *V. longisporum* hybridizations (D1 and D3) are phylogenetically closer related to

169 *V. dahliae* than parent A1 (39), *V. longisporum* genome alignments to *V. dahliae* display a
170 bimodal distribution with minima around 96.0% identity (Figure S2). To separate the two
171 sub-genomes, we assigned regions with an average sequence identity to *V. dahliae* smaller
172 than this minimum to parent A1 whereas regions with an identity larger than this threshold
173 were assigned to the D parent (Figure 1A). In this manner, 36.0-36.5 Mb were assigned to the
174 A1 parents and 34.7-35.9 Mb to the D parents (Table 1). Thus, the sub-genome sizes are quite
175 similar for each of the isolates and correspond to the expected genome sizes of haploid
176 *Verticillium* species (44, 45). Intriguingly, the majority of the *V. longisporum* chromosomes
177 is composed of DNA regions that originate from different parents, and only two
178 chromosomes have a single parental origin in each of the strains (Figure 1, Table S1). Thus,
179 *V. longisporum* chromosomes generally are mosaics of DNA regions of different parental
180 origin.

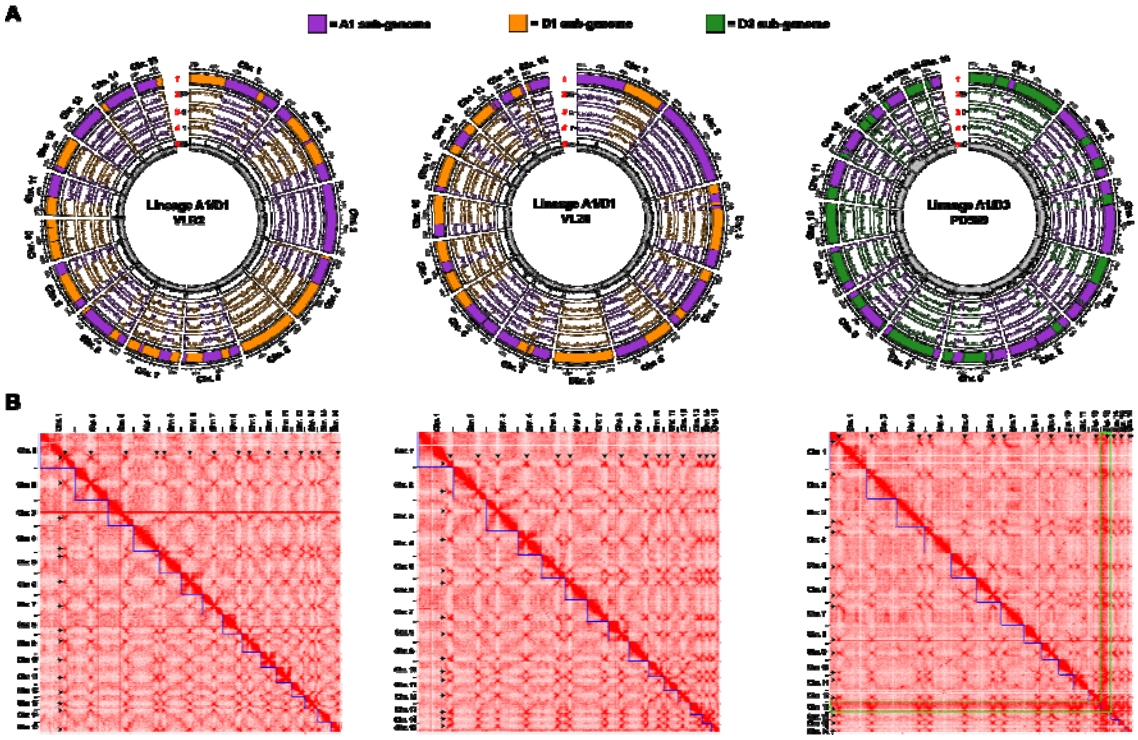


Figure 1. *Verticillium longisporum* displays a mosaic genome structure. (**A**) the *V. longisporum* chromosomes of strains VLB2, VL20 and PD589 are depicted. The different lanes in the circular plots represent: 1) regions assigned to Species A1 (purple), Species D1 (orange) and D3 (green); 2) sequence similarity of *V. longisporum* alignments to *V. dahliae* (% identity); 3) difference in sequence identity in percent point (pp) between exonic regions of *V. longisporum* double-copy genes. Only gene pairs with an ortholog in *V. dahliae* are depicted. Alleles with a higher identity to *V. dahliae* are depicted as a positive pp difference, whereas the corresponding homolog as a negative pp difference; 4) the relative difference in GC content (dGC) between genes in double copy; and 5) Read depth with non-overlapping windows of 10 kb. Data points of lanes 3-5 represent the average value of a window of eleven genes, which proceed with a step of one gene. (**B**) Hi-C contact frequency matrices for the three *V. longisporum* strains are shown. The red color indicates the contact intensity between genome region and the blue squares represent the pseudochromosomes. Centromeres display strong inter-chromosomal contacts and are visible as red dots outside the pseudochromosomes and are indicated with black arrows. Pseudochromosome 13 of strain PD589 generally displays stronger interactions than the other pseudochromosomes, which is outlined in green.

196 **Table 1. Comparison of *Verticillium longisporum* and *Verticillium dahliae* genome assemblies.**

	V. longisporum VLB2[1]	*V. longisporum* VL20[1]	*V. longisporum* PD589	*V. dahliae* JR2[2]
Genome size	72.7 Mb	72.2 Mb	72.0 Mb	36.2 Mb
Assigned to A1 sub-genome	36.2 Mb	36.5 Mb	36.0 Mb	/
Assigned to D sub-genome	35.9 Mb	35.1 Mb	34.7 Mb	/
Undetermined	0.6 Mb	0.6 Mb	1.3 Mb	/
Number of chromosomes	15	15	16/17[4]	8
Number of predicted genes	18,679	18,592	18,251	9,636
Assigned to A1 sub-genome	9,342	9,343	8,961	/
Assigned to D sub-genome	9,298	9,188	9,229	/
Undetermined	39	61	61	/
Number of predicted genes encoding secreted proteins	2,084	2,049	1,960	1,071
Assigned to A1 sub-genome	1,052	1,041	952	/
Assigned to D sub-genome	1,025	1,004	1000	/
Undetermined	7	4	8	/
Repeat content	14.55%	14.54%	12.78%	11.69%
BUSCO completeness[3]	99.1%	99.3%	97.9%	98.6%

197 [1]Previously published assemblies were reassembled using Hi-C sequencing (40)

198 [2](46)

199 [3]Based on Ascomycota Benchmarking Universal Single Copy Orthologs (BUSCOs)

200 [4]Total chromosome number is uncertain as PD589 contains one (partially) duplicated chromosome

The mitochondrial genome is inherited from the A1 parent in all lineages

To determine the phylogenetic position of the parental sub-genomes of the *V. longisporum* hybridization, we used the *V. longisporum* sub-genome sequences and previously published genome sequences of the haploid *Verticillium* species (45, 46) to construct a phylogenetic tree based on 1,520 Ascomycete benchmarking universal single copy orthologs (BUSCOs) that were present in a single copy in all analyzed *Verticillium* lineages. In accordance with previous phylogenetic studies (39, 40), the A1 parents diverged earlier from *V. dahliae* than the D1 and D3 parents (Figure S3). Furthermore, the D1 parent diverged earlier from *V. dahliae* than the D3 parent. We also constructed a phylogenetic tree based on mitochondrial DNA to determine the parental origin on the mitochondria. The *V. longisporum* mitochondrial genomes were assembled in a single contig with overlapping ends, indicating their circular nature. The mitochondrial genomes of the three *V. longisporum* strains were all 26.2 kb in size and were more than 99.9% identical in sequence. The phylogenetic position of the *V. longisporum* mitochondrial genomes clusters with the mitochondrial genomes of *V. alfalfae* and *V. nonalfalfae* (Figure S3). As the mitochondrial genome sequence is almost identical for three strains that are derived from the two hybridization events, the common A1 parent is the likely donor of the mitochondria.

Genomic rearrangements are responsible for the mosaic nuclear genome

Typically, a mosaic structure of a hybrid nuclear genome can originate from gene conversion or from chromosomal rearrangements between DNA strands of different parental origin (27). To analyze the extent of gene conversion, protein-coding genes were predicted for the *V. longisporum* strains using BRAKER with RNA-Seq data from fungal cultures grown *in vitro* (47). The number of predicted genes ranged from 18,251 to 18,679 for the different *V. longisporum* strains, which is 89-94% higher than the genes number of *V. dahliae* strain JR2

226 predicted using the same methodology (9,636 genes) (Table 1). In total, 8,961-9,343 genes
227 were assigned to the sub-genome of parent A1, whereas the number of genes in the D3 sub-
228 genomes ranged from 9,188 to 9,298 (Table 1). Thus, the gene numbers are similar for the
229 different *V. longisporum* sub-genomes and comparable to the gene number of *V. dahliae*.
230 Over 79% of the *V. longisporum* genes have one homolog, i.e. they occur in two copies,
231 which can originate from gene duplication (paralogy) or from the hybridization event
232 (homeology) (Figure 2). Within each of the *V. longisporum* sub-genomes, most genes (96.9-
233 99.6%) have no additional homolog and occur in a single copy (Figure 2B), indicating that
234 most homologous gene pairs in each *V. longisporum* genome are homeolog in nature and that
235 gene conversion only played a minor role after hybridization. To find traces of gene
236 conversion during their evolution, the sequence identity of 6,213 genes that have two
237 homologous copies in the two A1/D1 strains was compared, as these two strains belong to
238 distinct populations (40). Only eight genes were found to be highly similar (<1% nucleotide
239 sequence diversity) in VLB2, whereas the corresponding gene pair in VL20 was more diverse
240 (>1%) (Figure 3A). Similarly, in *V. longisporum* strain VL20, six highly similar copies were
241 found that are more divergent in VLB2, thereby confirming that gene conversion has hitherto
242 only played a marginal role in the evolutionary aftermath of the *V. longisporum*
243 hybridization.

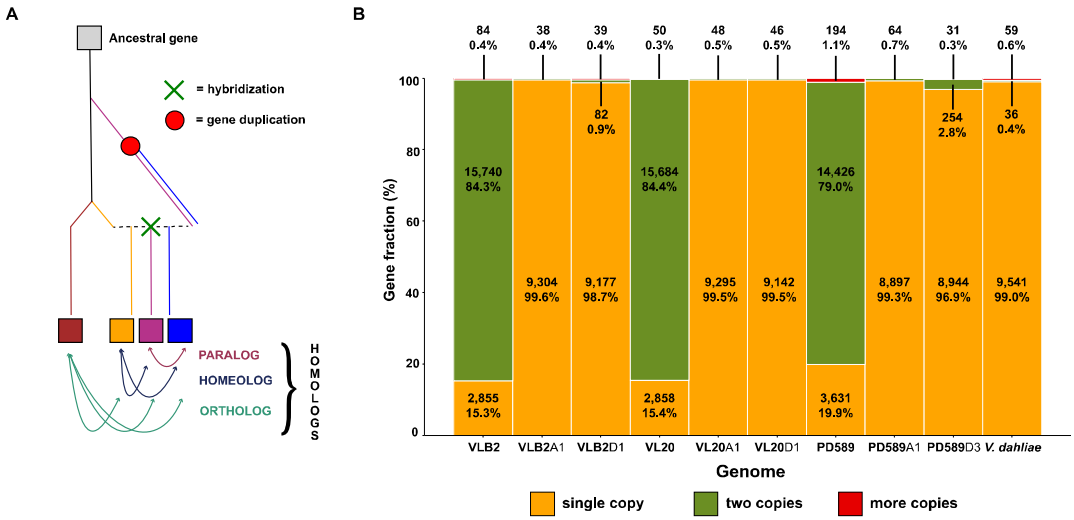


Figure 2. *V. longisporum* genes with two copies are almost exclusively homeologs. (**A**) Schematic overview of different evolutionary origins of homologous genes in hybrids. Paralogs are homologous genes that originate from gene duplication, while orthologous genes originate by speciation. Homeologs are homologous genes originating from a hybridization event. (**B**) The gene fraction occurring in single, two, and more than two copies in the *V. longisporum* strains VLB2, VL20 and PD589 are shown, with *V. dahliae* (strain JR2) as comparison. "A1", "D1" and "D3" represent species A1, D1 and D3 sub-genomes, respectively.

Considering that gene conversion played only a minor role during genome evolution (Figure 3), the mosaic genome structure of *V. longisporum* likely originated from rearrangements between homeologous chromosomes. To identify chromosomal rearrangements after the hybridization event that led to the A1/D1 lineage, the genome of *V. longisporum* strain VLB2 was aligned to that of strain VL20, revealing 24 syntenic breaks (Figure 3B). Rearrangement occurred in the majority of the chromosomes as only 2 and 3 chromosomes did not have syntenic breaks in VLB2 and VL20, respectively (Figure 3B). As genomic rearrangements are often associated with repeat-rich genome regions, such as in *V. dahliae* (46, 48, 49), the synteny break points were tested for their association to repetitive regions. Since the median repeat fraction in a 20 kb window around the repeats is 15.5%, which is significantly more than the median repeat fraction based on random sampling

(average = 3.4%, σ = 1.7%) (Figure S4), it can be concluded that the chromosomal rearrangements are similarly associated with repeats also in *V. longisporum*. In conclusion, chromosomal rearrangement rather than gene conversion is the main mechanism explaining the mosaic structure of the *V. longisporum* genome.

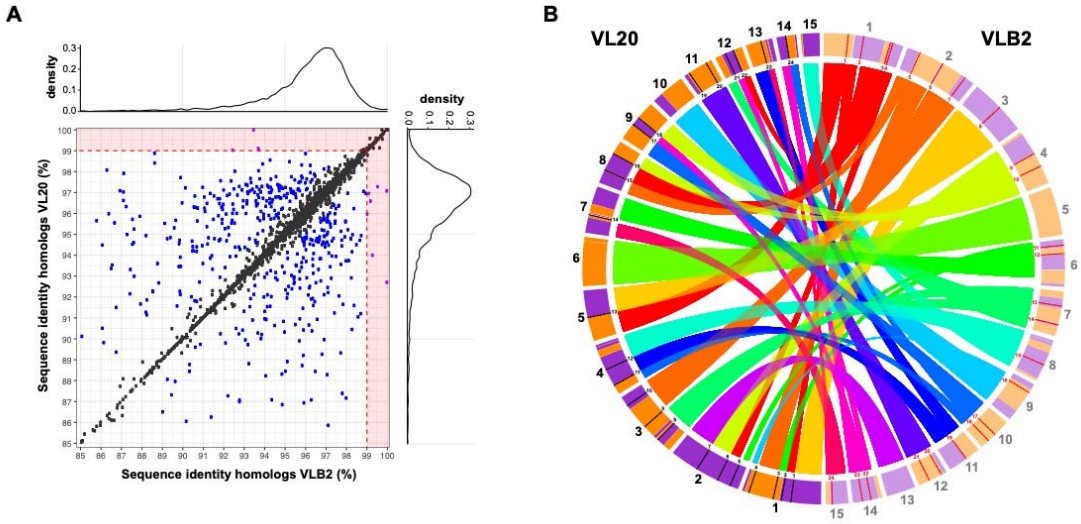


Figure 3. The mosaic genome structure of *Verticillium longisporum* originates from genomic rearrangements. (**A**) The contribution of gene conversion to *V. longisporum* genome evolution. Sequence identities between genes in copy, present in *V. longisporum* VLB2 and VL20, are depicted. Homologous gene pairs within a strain that encountered gene conversion are expected to have higher similarity within a strain compared with the corresponding gene pair in the other strain. Gene pairs with divergence of more than one percent in one *V. longisporum* strain and less than one percent in the other strain were considered conserved in the latter strain (purple dots in the red zones). In other cases, pairs that differ less than one percent are depicted as a black dot, whereas a difference greater than one percent is depicted as a blue dot. (**B**) The contribution of genomic rearrangements to *V. longisporum* genome evolution. The *V. longisporum* chromosomes of strains VLB2 (displayed on the right) and VL20 (displayed on the left) are depicted. Ribbons indicate syntenic genome regions between the two strains and contig colors indicate the parental origin similar to Figure 1 (purple = A1 and orange = D1). The red and black lines with the associated numbers on the chromosomes indicate synteny breaks.

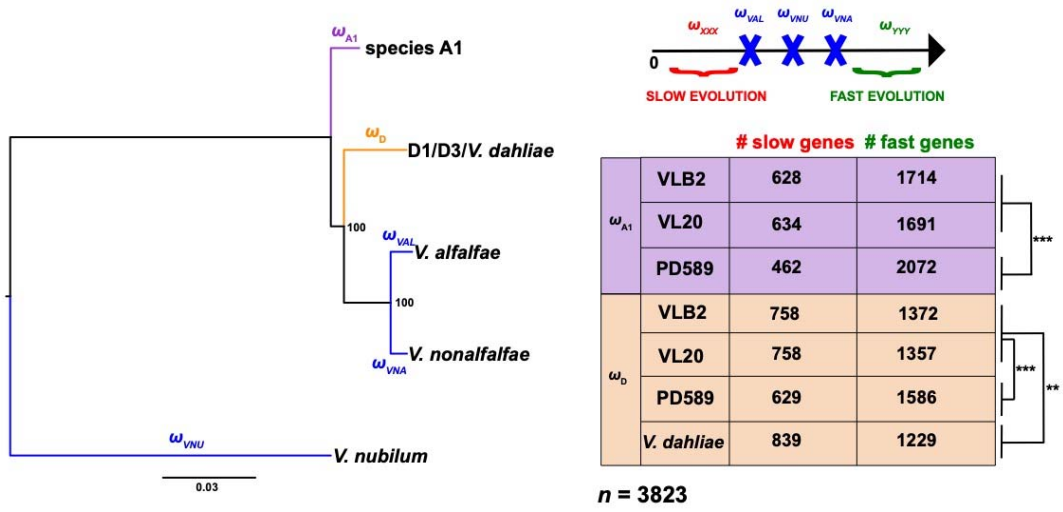
V. longisporum loses heterozygosity through deletions

To study putative gene losses in the aftermath of hybridization, we determined genes that have no homeolog or paralog, and can thus be considered to occur in single copy. For the A1/D1 isolates, 15.3-15.4% of the genes occur in single copy, whereas this is 19.9% for A1/D3 isolate PD589 (Figure 2B). We checked if proteins encoded by single copy genes are enriched for particular Gene Ontology (GO) terms, Clusters of Orthologous Groups (COGs), or encoding a protein with a signal peptide, which suggests that these proteins are secreted. No GO terms and COGs were enriched for the single copy genes in any of the *V. longisporum* strains (Fisher's exact test with Benjamini-Hochberg correction, p-value < 0.05). In total, 7.8-10.2% of the single copy genes encodes a protein with a signal peptide, which is a significantly lower than the 11.9-12.3% for genes with a homologous copy in the same genome (Fisher's exact test, p-value < 0.05). Of the A1/D1 single copy genes, 52% reside in the A1 sub-genome and 47% in the D1 sub-genome. Similarly, for PD589, 49% and 50% reside in the A1 and D3 sub-genome, respectively. Thus, single copy genes are equally distributed across the two sub-genomes in *V. longisporum*. Single copy genes can either originate from gene loss or from parent-specific contributions to the hybrid. Since VLB2 and VL20 originate from the same hybridization event (40), we can quantify how many single copy genes originate from gene loss during divergence of VLB2 and VL20. In total, 14.7-14.8% of the singly copy genes have at least one copy in each sub-genome of the other A1/D1 strain, suggesting that gene deletion is an on-going process in *V. longisporum* evolution. Of the single copy genes that lost their homeolog after the hybridization event, 48% resided in the species A1 sub-genome, whereas 51-52% in the D1 sub-genome, suggesting that gene losses occurred to a similar extent in each of the sub-genomes.

304 **Acceleration of gene evolution upon hybridization**

305 To investigate the gene sequence evolution subsequent to hybridization, we compared the
306 ratio of non-synonymous (*Ka*) and synonymous (*Ks*) substitutions (ω) for branches leading to
307 *Verticillium* species (Figure 4). To exclude the putative impact of the (partial) chromosome
308 13 duplication in PD589, we excluded genes of this chromosome from the analysis.
309 Substitution rates were determined for a total of 3,823 genes that have just one ortholog in the
310 analyzed *Verticillium* species, *V. alfalfae*, *V dahliae, V. nonalfalfae* and *V. nubilum*, as well
311 as in each of the *V. longisporum* sub-genomes. To mitigate possible biases of different
312 divergence times between the *Verticillium* species, we performed the analyses four times:
313 three times with the two sub-genomes of *V. longisporum* strains VLB2, VL20, and PD589,
314 and once with *V. dahliae* and the A1 sub-genome of VLB2 (Figure 4). *V. longisporum* and *V.*
315 *dahliae* genes with higher ω than their *V. alfalfae*, *V. nonalfalfae* and *V. nubilum* orthologs
316 were considered quickly evolving, whereas those with lower ω were considered slowly
317 evolving. Comparing the D1/D3/*V. dahliae* branch, *V. dahliae* has 839 slowly evolving
318 genes, which is a higher number than the 758 and 629 slowly evolving genes of the *V.*
319 *longisporum* D1 and D3 sub-genomes, respectively. Conversely, *V. dahliae* has 1,229 quickly
320 evolving genes, which is lower than the number found for the *V. longisporum* D1 and D3
321 sub-genomes, 1,357/1,372 (VL20/VLB2) and 1,586, respectively (Figure 4). This observation
322 fits to the prevailing hypothesis that hybridization accompanied by genome duplication has a
323 'relaxing' effect on gene evolution (32, 50). Furthermore, the lower number of slowly
324 evolving genes and larger number of quickly evolving genes in the D3 sub-genome is
325 significantly different from the D1 sub-genome (Fisher's exact test, $P < 0.001$). Similar to the
326 D sub-genomes, the A1 sub-genome of lineage A1/D3 has higher number of quickly evolving
327 genes (2,072 vs. 1,691-1,714) and lower number of slowly evolving genes (462 vs. 628-634)
328 than the A1 sub-genome of lineage A1/D1. In conclusion, *V. longisporum* lineage A1/D3

329 genes generally evolve faster than lineage A1/D1 genes in both sub-genomes. This may
330 indicate that A1/D3 evolved a longer time under the more relaxes gene evolutionary
331 conditions than A1/D1, i.e. A1 and D3 hybridized a longer time ago than A1/D1.



332

333 **Figure 4. *Verticillium longisporum* genes divergence faster than *Verticillium dahliae* orthologs.** *Ka/Ks* ratios
334 (ω) were calculated for the tree branches leading to *Verticillium* spp. of the clade Flavnonexudans genomes and
335 the *V. longisporum* sub-genomes. A total of 3,823 genes with one ortholog in all respective *Verticillium* (sub-
336)genomes were analyzed. *V. longisporum* and *V. dahliae* genes with fast or slow evolution have a higher ω or
337 lower ω, respectively, than their *V. alfalfae*, *V. nonalfalfae* and *V. nubilum* orthologs. Significance in gene
338 numbers was calculated with the Fisher's exact test. **: $P < 0.01$ and ***: $P < 0.001$.

339

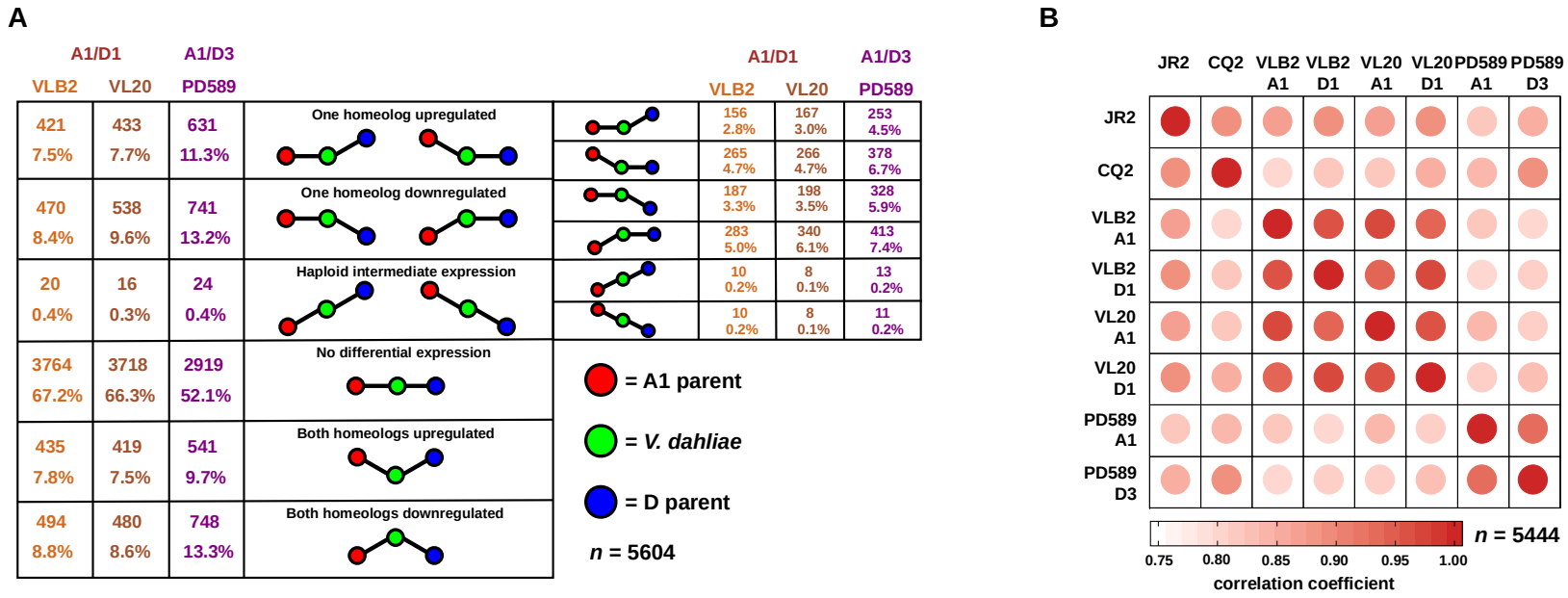
340 To see whether particular genes evolve faster, we functionally characterized the *V.*
341 *longisporum* A1/D3 genes that have a higher ω than their *V. alfalfae*, *V. nonalfalfae* and *V.*
342 *nubilum* orthologs, but also higher than their lineage A1/D1 homologs from the
343 corresponding A1 and D sub-genomes to select genes that quickly evolved after the A1 and
344 D1/D3 last common ancestor. In total, 1,350 of the 3,823 (35.3%) analyzed genes were
345 quickly evolving in PD589 A1 sub-genome and 1,084 (28.4%) for the D3 sub-genome. We
346 screened for GO term, COG and secreted protein enrichments in these fast evolving A1/D3
347 genes and no enrichments for the COGs and for genes encoding secreted proteins were found.

In the A1 sub-genome 3 GO terms with a molecular function were significantly enriched, associated with molecule binding (protein and ATP) and ATPase activity. In the D3 sub-genome, "ATP binding" was the only significantly enriched GO term, which was also enriched in the A1 sub-genome. In conclusion, the more pronounced "gene relaxation" in the A1/D3 lineage when compared with the A1/D1 lineage does not clearly seem to affect genes with particular functions.

Expression pattern homogenization in the hybridization aftermath

To investigate the impact of hybridization on gene expression, the expression of *V. longisporum* genes was compared with *V. dahliae* orthologs from strains grown *in vitro* in potato dextrose broth. To this end, expression of single copy *V. dahliae* genes was compared with orthologs that are present in two homeologous copies in three *V. longisporum* strains (VLB2, VL20, and PD589). Genes on chromosome 13 from strain PD589 and their homologs were excluded from the analysis to avoid putative biases due to a (partial) chromosome duplication, and in total 5,604 expressed genes were compared. RNA sequencing reads were mapped to the predicted *V. longisporum* genes of which 50-51% mapped to species A1 homeologs and 49-50% to the D homeologs. Thus, we observed no global differences in overall contribution to gene expression of the sub-genomes. Over half of the *V. longisporum* homeologs display no differential expression with their *V. dahliae* ortholog, indicating that the majority of the genes did not evolve differential expression patterns (Figure 5A). In both lineages, higher numbers of differently expressed genes were found in the A1 sub-genome than in the D sub-genomes; 27 vs. 23% for A1/D1 and 38 vs. 34% for A1/D3, respectively. The higher fraction of differentially expressed A1 genes is in accordance with the more distant phylogenetic relationship of parent A1 with *V. dahliae* than of the D parents (Figure S3). Intriguingly, although D3 diverged more recently from *V. dahliae* than D1, D3 has more

differentially expressed orthologs with *V. dahliae* than D1. When comparing expression patterns between sub-genomes, 11-13% of the genes display differential expression between their A1 and D homeologs. Intriguingly, this is more than half the number of differentially expressed D and *V. dahliae* orthologs (23-34%), despite the fact that the D parents diverged more recently from *V. dahliae* than from species A1 (Figure S3). In general, the gene expression patterns of the A1 and D sub-genomes of the same hybridization event are highly correlated (0.93-0.96), higher than D sub-genomes and *V. dahliae* strain JR2 (0.85-0.89) and higher than the A1 sub-genomes between hybridization events (0.82-0.84) (Figure 5B; Table S2). To compare these expression patterns with the gene expression variation between different *V. dahliae* strains, we sequenced RNA from the cotton-infecting *V. dahliae* strain CQ2 grown in potato dextrose broth. Although JR2 and CQ2 belong to the same species, their overall gene expression pattern is more dissimilar (ρ = 0.89) than that of *V. longisporum* sub-genomes (Figure 5B; Table S2). The overall discrepancy in phylogenetic relationship and expression pattern similarities suggests that sub-genome expression patterns of the sub-genomes in *V. longisporum* homogenized upon hybridization.



388

389 **Figure 5. Gene expression patterns of *Verticillium longisporum* sub-genomes display remarkable resemblance.** Expression pattern comparison between *Verticillium*
390 *longisporum* sub-genomes and *Verticillium dahliae* in culture medium (**A**) Differential expression between *V. longisporum* and *V. dahliae* genes. Only genes with one
391 homolog in *V. dahliae* and two homeologs in the *V. longisporum* strains VLB2, VL20 and PD589 were considered for differential expression. The significance of differential
392 expression was calculated using *t*-tests relative to a threshold of log2-fold-change of 1 and a Benjamini-Hochberg corrected *p*-value cut-off of 0.05. (**B**) Expression pattern
393 correlation between *V. longisporum* and *V. dahliae*. Only genes with one homolog in *V. dahliae* strains JR2 and CQ2 and two homeologs in the *V. longisporum* strains VLB2,
394 VL20 and PD589 were considered. Spearman's correlation coefficients (ρ) were calculated based on the mean transcripts per million values of three replicates.

Differential homeolog expression occurs in particular gene categories

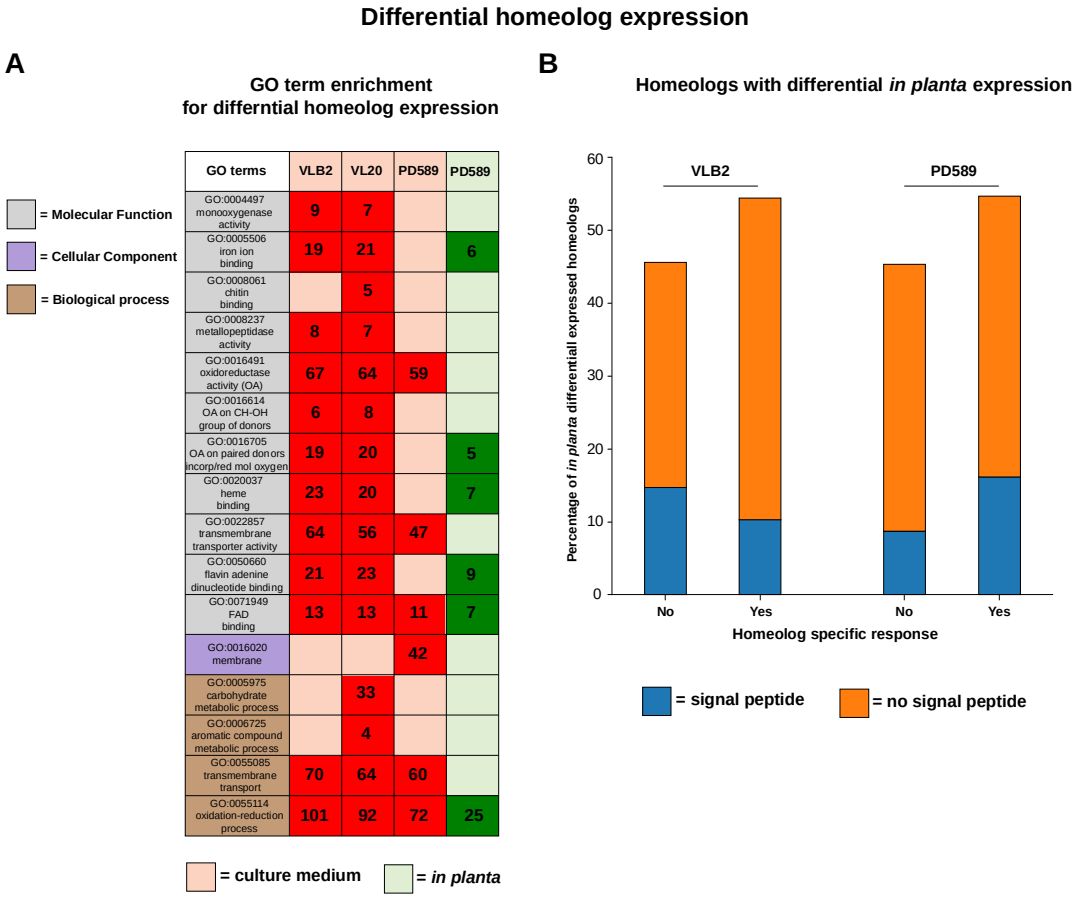
Although parental gene expression patterns appear to have globally homogenized upon hybridization, differential homeolog expression occurs as well (Figure 5). To assess if genes with differential homeolog expression belong to specific gene groups, we screened for functional enrichments. In total, 10% of the fast-evolving PD589 genes (defined in the previous section) have different homeolog expression, which is significantly lower than the 12% different homeolog expression for the remainder of the genes (Fisher's exact test, $P < 0.05$). In both A1/D1 and A1/D3 lineages, genes with differential homeolog expression are enriched for GO terms related to oxidation-reduction processes, transmembrane transport and FAD binding (Figure 6A, Table S3). Additionally, the COGs "carbohydrate transport and metabolism" and "secondary metabolites biosynthesis, transport, and catabolism" (Q) are enriched in both lineages (Table S3). Furthermore, we tested if genes encoding secreted proteins were significantly enriched among the genes with differential homeolog expression. Indeed, 23 and 16 % of the genes with different homeolog expression code for a secreted protein in the lineage A1/D1 isolates and in the A1/D3 isolate, respectively, whereas this is 9% of the genes that do not display differential expression among homeologs (VLB2 P =1.23E-32, VL20 P =3.71E-29 and PD589 P =1.14E-08, Fisher's exact test). In conclusion, differential homeolog expression seems to be important for particular gene categories, including categories that can be implicated in plant pathogenicity.

Homeolog-specific expression responses upon plant colonization

Considering the plant pathogenic nature of *V. longisporum*, and also that genes encoding secreted proteins, which are often implicated in pathogenicity on host plants, are enriched among the genes with differential homeolog expression, we assessed homeolog-specific gene expression during plant colonization. To this end, oilseed rape plants were inoculated with

the *V. longisporum* strains VLB2, VL20, and PD589. As observed previously, oilseed rape plants inoculated with VLB2 and PD589 developed typical *Verticillium* symptoms including stunted plant growth and leaf chlorosis (51). In contrast, oilseed rape plants inoculated with VL20 did not display any disease symptoms. Consequently, we performed total RNA sequencing for oilseed rape plants inoculated with *V. longisporum* strains VLB2 and PD589. For strain PD589, genes on chromosome 13 and their homeologs were removed from the analysis. For VLB2 and PD589, 51% of the reads mapped to the A1 sub-genome and 49 to the D sub-genome. Thus, similar to *in vitro* grown *V. longisporum*, we did not observe any global difference in overall contribution to gene expression of one of the sub-genomes *in planta*. In total, 1.1% and 2.7% of the homeologs displayed differential expression *in planta*, which is less than the 11.3 and 13.4% found for VLB2 and PD589 grown *in vitro*, respectively. Genes with differential homeolog expression *in planta* were not enriched for any GO term in the A1/D1 strain VLB2 (Table S3), whereas in the A1/D3 strain PD589, differentially expressed homeologs were enriched for GO terms associated with oxidation-reduction processes and molecular binding (iron ion, heme, flavin adenine dinucleotide and FAD) (Figure 6A, Table S3). For A1/D1 and A1/D3, genes with different homeolog expression were enriched for those encoding secreted proteins; 25% of differentially expressed homeologs encode secreted proteins and 8-9% of the non-differentially expressed homeologs encode other proteins ($P < 0.05$, Fisher's exact test). Thus, similar to *in vitro V. longisporum* grown, differential homeolog expression *in planta* is especially important for genes encoding secreted proteins. In 33% of these secretome genes with differential homeolog expression *in planta*, no Pfam domain could be annotated, which is a feature often observed for effector proteins as they are often examples of biological innovation (52). Of these genes that could be functionally annotated, a carbohydrate-active enzyme (CAZyme) function was annotated in 32% of the cases. The remaining part of the functionally annotated

445 genes with differential homeolog expression included other enzymes, such as proteases,
446 lipases, carboxylesterases and peroxidases. We compared genes with differential homeolog
447 expression *in planta* and *in vitro* to assess potential correlation. Intriguingly, over half (54-
448 55%) of the differentially expressed homeologs *in planta* are not differentially expressed in
449 culture medium or have the inverse expression pattern, e.g. *in vitro*: A1>D and *in planta*
450 A1<D (Figure 6B). Thus, over half of the genes with a differential homeolog expression *in*
451 *planta* display a homeolog-specific response compared to *in vitro* growth. For VLB2, 19% of
452 these genes with a homeolog-specific response encode secreted proteins, whereas 32% of
453 genes with similar differential homeolog expression *in planta* and *in vitro* encode secreted
454 proteins. The opposite pattern in observed for PD589, i.e. 30% with a homeolog-specific
455 response and 19% with similar differential homeolog expression *in planta* and *in vitro*.
456 However, these differences were not significant ($P > 0.05$, Fisher's exact test). In conclusion,
457 different growing conditions cause homeolog-specific changes in the majority of the *V.*
458 *longisporum* genes with differential homeolog expression, which is enriched in genes that
459 encode secreted proteins.



460

461 **Figure 6. *Verticillium longisporum* displays sub-genome-specific gene expression responses.** Functional
462 enrichments for *Verticillium longisporum* genes with different homeolog expression in culture medium and *in*
463 *planta*. Only *V. longisporum* genes with two homeologs were considered. (**A**) Gene Ontology (GO) terms that
464 are significantly enriched in differentially expressed homeologs of VLB2, VL20 and PD589 are displayed. A
465 more detailed overview and level of significance are reported in Table S3. The number of genes with differential
466 homeolog expression are indicated. (**B**) Fractions of genes with differential homeolog expression *in planta* with
467 and without a homeolog specific response. Genes have a homeolog specific response as they display differential
468 homeolog expression *in planta* and have no differential or the opposite expression ratio for *V. longisporum*
469 grown in culture medium.

Discussion

Hybridization is a powerful evolutionary mechanism that can lead to the emergence of new plant pathogens with distinct features when compared with their parents (8, 23). Here, we reveal the transcriptomic plasticity of the hybrid pathogen *V. longisporum* and illustrate the parental allele-specific response to different environmental cues. Differential expressed *V. longisporum* homeologs are enriched for genes encoding secreted proteins that generally act to facilitate environmental manipulation (53). Interestingly, over half of the differentially homeolog expressed genes *in planta* display different relative contributions *in vitro*. Thus, upon the environmental changes that are associated with different growth conditions, *V. longisporum* encounters sub-genome specific gene expression alterations, leading to differential homeolog expression. Although not previously reported for any other hybrid plant pathogen, sub-genome specific gene expression alterations has previously been reported to occur in the artificial yeast hybrid *S. cerevisiae* x *Saccharomyces uvarum* upon temperature change (54). Genes with these sub-genome specific responses were involved in a variety of biological processes, including the trehalose metabolic process that is involved in thermotolerance. Thus, more generally, hybrid fungi, comprising natural as well as artificial hybrids, respond to environmental change in an allele-specific manner, especially for genes that manipulate or mitigate environmental changes. Secretome genes with a differential homeolog expression *in planta* often have an enzymatic function or lack an annotated Pfam domain, which is a feature often observed for effector proteins that act in pathogenicity (52). Thus, conceivably, homeolog specific responses *in planta* occurs in genes that are important for host colonization. Similarly, differential homeolog expression in the hybrid opportunistic human pathogen *Candida orthopsilosis* involves genes that are implicated in host interactions, related to superoxide dismutase activity and zinc metabolism (55).

494 Although differential homeolog expression occurs, the general tendency is that
495 expression patterns between the A1 and D sub-genomes homogenized upon hybridization
496 (Figure 5). Despite the absence of A1 and D1 species due to their enigmatic nature, we can
497 conclude that parental gene expression patterns homogenized in the aftermath of
498 hybridization as sub-genome expression patterns display more resemblance than the
499 expression pattern between *V. dahliae* and the D parents and between the A1 sub-genomes of
500 different hybridization events (Figure 5B; Table S2). Homogenization of parental expression
501 patterns has been similarly observed in the fungal allopolyploid *Epichloë* Lp1 (36) as well as
502 in the artificial hybrid *S. cerevisiae* x *S. uvarum* where the extent of differential ortholog
503 expression between the parents was diminished upon hybridization (56). Thus, gene
504 expression homogenization seems to be a more general phenomenon in fungi. Gene
505 expression divergences may evolve through mutations in regulatory sequences of the gene
506 itself (*cis*-effects), such as promoter elements, or alterations in other regulatory factors (*trans*-
507 effects), such as chromatin regulation (57, 58). Conceivably, the higher correlation in
508 homeolog expression patterns than parental ortholog expression patterns originates from
509 changes in *trans* regulators as homeologs, in contrast to orthologs, share the same nuclear
510 environment (58). Intriguingly, parent D3 has more genes that are differentially expressed to
511 *V. dahliae* orthologs than parent D1, despite that D3 diverged more recently from *V. dahliae*
512 than D1 (Figure 5, Figure S4). Correspondingly, the A1 sub-genome of the lineage A1/D3
513 displays more differential gene expression with *V. dahliae* than the A1 sub-genome of the
514 A1/D1 lineage. This can indicate that A1 and D3 hybridized before A1 and D1, as more
515 distinct expression patterns may have evolved over time.

516 In addition to the transcriptomic plasticity of homeolog expression upon
517 environmental changes, *V. longisporum* is also plastic on a genomic level, which is displayed
518 by its mosaic structure (Figure 1A, Table S1). Mosaicism is also observed in the grass

519 pathogen *Zymoseptoria pseudotritici*, which is a close relative of the prominent wheat
520 pathogen *Zymoseptoria tritici* (29). *Z. pseudotritici* is a homoploid hybrid that displays
521 mosaicism on a population level where genome regions inherited from one parent display low
522 variation, whereas high variable genome regions were transmitted from both parents. *V.*
523 *longisporum* mosaicism is caused by extensive genomic rearrangements after hybridization
524 (Figure 2B, 3). Genomic rearrangements are major drivers of evolution and facilitate
525 adaptation to novel or changing environments (48). Genomic rearrangements are not specific
526 to the hybrid nature of *V. longisporum* as other *Verticillium* species similarly encountered
527 extensive chromosomal reshuffling (44, 45, 49, 59). In *V. dahliae*, genomic rearrangements
528 especially occur in genomic regions that were originally described as lineage-specific
529 regions, which are enriched for active transposable elements, and that are derived from
530 segmental duplications that were followed by extensive reciprocal gene losses, encounter
531 nucleotide sequence conservation and have an unique epigenomic profile (49, 59–62). These
532 lineage-specific regions are enriched for *in planta* expressed genes and contain effector genes
533 that facilitate host infection (59, 60, 63, 64). Since more recently, these lineage-specific
534 regions are referred to as dynamic chromosomal regions (60). Similar to *V. dahliae*, syntenic
535 breaks in *V. longisporum* often reside in repeat-rich genome regions as repetitive sequences
536 (Figure S3), due to their abundance, are more likely to act as a substrate for unfaithful repair
537 of double-strand DNA breaks (48, 49). However, the presence of two genomes within a
538 single hybrid nucleus may also provide homeologous sequences with sufficient identity to
539 mediate unfaithful repair.

540 The *V. longisporum* D genomes globally display accelerated evolution when
541 compared with their *V. dahliae* orthologs (Figure 4), which may be a consequence of genome
542 doubling. Interestingly, the *V. longisporum* A1/D3 lineage strain PD589 encountered a more
543 divergent gene evolution when compared with the A1/D1 lineage strains VLB2 and VL20 in

544 both sub-genomes, indicating that the A1/D3 hybridization occurred prior to the A1/D1
545 hybridization as a longer allodiploid state could facilitate extended sequence divergence (65).
546 However, accelerated evolution is not consistently observed in fungi as deceleration upon
547 allopolyploidization has been recorded in the fungal genus *Trichosporon* (66). Arguably,
548 environmental cues play an important role in the speed and grade of gene diversification upon
549 allopolyploidization (67). Possibly, accelerated gene evolution in *V. longisporum* is cued by a
550 host range alteration as it is, in contrast to haploid *Verticillium* species, a Brassicaceae
551 specialist (42). However, we did not find functional enrichments in fast evolving genes that
552 points towards that hypothesis. Moreover, as the A1 species remains enigmatic, we cannot be
553 sure a host shift occurred (39, 41).

554 Whole-genome duplication events are typically followed by extensive gene loss, often
555 leading to reversion to the original ploidy state (68). For instance, the artificial interspecific
556 hybrid *S. cerevisiae* x *S. uvarum* encountered nine independent events where loss of
557 heterozygosity occurred after evolving for hundreds of generations under nutrient-limited
558 conditions (69). Heterozygosity loss has only proceeded to a limited extent in *V.*
559 *longisporum*, as 84% of lineage A1/D1 genes and 79% of lineage A1/D3 genes are present in
560 two copies, whereas the haploid *V. dahliae* only contains 0.4% of its genes in two copies
561 (Figure 2B). Thus, the *V. longisporum* genome displays the symptoms of a recent allodiploid,
562 with gene loss being an on-going process that by now has only progressed marginally.
563 Heterozygosity loss can indicate deleterious epistatic interactions between parental genomes
564 that need to homogenize in order for the hybrid to be viable. Similar to other fungal hybrids
565 (69, 70), we did not observe a specific group of genes where loss of heterozygosity was
566 selected for. The degree of haploidization is a third indication that the A1/D3 lineage likely
567 hybridized prior to A1/D1, as haploidization progress further in A1/D3 than in A1/D1 (Figure
568 2B). *C. orthopsilosis* hybrids from different hybridization events have different degrees of

heterozygosity loss but genes were homeologs are maintained in both hybrids are enriched for those to have differential homeolog expression (55). Although species often revert to their original ploidy state after polyploidization, a retention of both homeolog copies can also be evolutionary advantageous, for instance to respond in a parental allele-specific fashion to environmental cues (Figure 6).

Conclusion

Allodiploidization is an intrusive evolutionary mechanism in fungi where two chromosome sets from parents with a distinct evolutionary history merge. Consequently, most genes obtain an additional gene copy that can be differentially regulated according to the environmental conditions. Thus, allodiploid fungi can respond in a parental allele-specific fashion to environmental cues. Besides such parental allele-specific gene expression, allodiploidization furthermore contributed to a dynamic genome evolution through rearrangements between parental chromosome sets and accelerated gene evolution in *V. longisporum*. Thus, in comparison to haploid *Verticillium* species, *V. longisporum* has a high adaptive potential that can contribute to host immunity evasion and may explain its specialization towards Brassicaceous plant hosts.

Material and Methods

V. longisporum genome sequencing and assembly

Genome assemblies of the *V. longisporum* strains VLB2 and VL20 were previously constructed using long reads obtained through single-molecule real-time (SMRT) sequencing (40). Here, we sequenced *V. longisporum* strains PD589 using Oxford Nanopore Technology (ONT). In order to obtain DNA of PD589, spores were harvest from PDA plates and grown in 1/5 PDB for 5 days. Mycelium and spores were collected on Myra cloth, freeze-dried overnight and ground to fine powder. For DNA isolation, 100 mg of material was used and incubated for one hour at 65°C with 800 µL DNA extraction buffer (0.35 M Sorbitol, 0.1 M Tris-base, 5 mM EDTA pH 7.5), nucleic lysis buffer (0.2 M Tris, 0.05 M EDTA, 2 M NaCl, 2% CTAB) and Sarkosyl (10% w/v) in a 2:2:1 ratio. Subsequently, ½ volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added, shaken vigorously and incubated at room temperature (RT) for 5 minutes before centrifugation at maximum speed (16,000 rpm) for 15 minutes (RT). The upper (aqueous phase) layer was transferred to a new tube, 5 µL of RNAase (10 mg/µL) was added and incubated at 37°C for one hour. Next, ½ volume of chloroform was added, mixed and centrifuged at maximum speed for 10 minutes at RT. The upper layer was transferred to a new tube and a second chloroform wash step was performed. After transferring the upper layer to a new tube, it was mixed with 1 volume (~ 800 µL) of 100% ice-cold ethanol by gently inverting the tube and finally the DNA was fished out and washed twice by applying 500 µL of 70% ethanol. Finally, the DNA was air-dried, resuspended in nuclease-free water and stored at 4°C overnight. The DNA quality, size and quantity were assessed by nanodrop, gel electrophoresis and Qubit analyses, respectively.

To sequence the *V. longisporum* strain PD589 DNA, a library was prepared as described in the manufactures protocol provided by ONT (SQK-RAD004) with an initial amount of ~ 400 ng HMW DNA. The library was loaded onto a R9.4.1 flow cell which ran

611 for 24 hours and yielded ~7 Gb of data. ONT sequencing reads were basecalled using Guppy
612 (version 3.1.5) using the high accuracy base calling algorithm. Subsequently, adapter
613 sequences were identified and removed using Porechop (version 0.2.3; default settings);
614 adapters at the end of the reads were trimmed and reads with internal adapters were
615 discarded. To be able to polish the genome assembly, we used the same HWA DNA isolated
616 for ONT sequencing to generate ~35 million high-quality (95% > phred score of 20) 150 bp
617 paired-end reads (~76x coverage) using the BGISeq platform (BGI Tech Solutions,
618 Hongkong).

619 The *V. longisporum* PD589 genome was *de novo* assembled using Canu (version 1.8;
620 genomeSize=70m, corOutCoverage=100, batOptions='-dg 3 -db 3 -dr 1 -ca 500 -cp 50') (71).
621 In total, 924,740 cleaned ONT reads were used for the *de novo* assembly of which 743,753
622 where >1 kb (~88x coverage). The genome assembly was polished using two sequential
623 rounds of Apollo (version 1.1) (72). To this end, the high-quality paired-end reads were
624 mapped to the genome assembly using bwa (version 0.7.17-r1188; default settings) (73).

625 To improve the assemblies to (near) chromosome level, chromatin conformation
626 capture (Hi-C) followed by high-throughput sequencing was performed for VLB2, VL20 and
627 PD589, similar as previously reported (44). For the three *V. longisporum* strains, one million
628 spores were added to 400 ml Potato Dextrose Broth and incubated for 6 days at 22°C with
629 continuous shaking at 120 rpm. 300 mg (fresh weight) mycelium was used as input for
630 generating Hi-C sequencing libraries with the Proximo Hi-C kit (Microbe) (Phase Genomics,
631 Seattle, WA, USA), according to manufacturer instructions. Hi-C sequencing libraries were
632 paired-end (2x150 bp) sequenced on the NextSeq500 platform at USEQ (Utrecht, the
633 Netherlands). Juicer (v1.6) was then used to map Hi-C sequencing reads to the previously
634 obtained assemblies (74). The contact matrices generated by Juicer were used by the 3D *de*
635 *novo* assembly (3D-DNA) pipeline (v180922) to eliminate misjoints in the previous

636 assemblies (75). The assemblies were manually further improved using Juicebox Assembly
637 Tools (JBAT) (v.1.11.08) (76). JBAT was subsequently used to determine centromere
638 location based on intra- and inter-chromosomal contact frequencies. Only contigs that were
639 larger than 100 kb were maintained in the assembly. Coverage of the ONT for the *V.*
640 *longisporum* PD589 assembly was determined for 20 kb windows with samtools depth (v1.9)
641 (77) and reads were mapped with minimap2 (v2.17-r941) (78).

642 The mitochondrial genomes of the haploid *Verticillium* species were previously
643 sequenced and assembled (45). Mitochondrial *V. longisporum* genomes were assembled
644 alongside the nuclear genomes (40). Mitochondrial contigs consisted of multiple copies of the
645 mitochondrial genome due to its circular nature. A single copy of the mitochondrial genome
646 was excised using BEDTools getfasta (v2.23.0) (Quinlan and Hall 2010). Filtered *V.*
647 *longisporum* sub-reads were mapped to these single-copy mitochondrial assemblies using
648 circlator (v1.5.5) (Hunt et al. 2015). The mapped reads were subsequently used to make a
649 new *V. longisporum* mitochondrial genome assembly using SAMtools mpileup (v1.8) (Li et
650 al. 2009).

651

652 **RNA sequencing**

653 To obtain RNA-seq data for *Verticillium* grown in culture medium, *V. dahliae* isolates JR2
654 and CQ2, and *V. longisporum* isolates VLB2, VL20 and PD589 were grown for three days in
655 potato dextrose broth (PDB) with three biological replicates for each isolate. To obtain RNA-
656 seq data from *in planta* growth, two-week-old plants of the susceptible oilseed rape cultivar
657 'Quartz' were inoculated by dipping the roots for 10 minutes in $1x10^6$ conidiospores ml^{-1}
658 spore suspension of *V. longisporum* isolates VLB2, VL20 and PD589, respectively (51).
659 After root inoculation, plants were grown in individual pots in a greenhouse under a cycle of
660 16 h of light and 8 h of darkness, with temperatures maintained between 20°C and 22°C

during the day and a minimum of 15°C overnight. Three pooled samples (10 plants per sample) of stem fragments (3 cm) were used for total RNA extraction. Total RNA was extracted based on TRIzol RNA extraction (Simms et al. 1993). cDNA synthesis, library preparation (TruSeq RNA-Seq short-insert library), and Illumina sequencing (single-end 50 bp) was performed at the Beijing Genome Institute (BGI, Hong Kong, China).

Gene prediction and functional characterization

The *V. longisporum* assemblies of strains VLB2, VL20 and PD589 and the previously published assemblies of *V. dahliae* strains JR2 and CQ2 (46, 61) were annotated using the BRAKER v2.1.4 pipeline with RNA-Seq data with the options "--softmasking" and "--fungus" enabled (47). RNA-seq reads from *Verticillium* grown in axenic culture (all replicates) were mapped to the assemblies using TopHat v2.1.1 (79). Predicted genes with internal stop codons, without a start codon or with an unknown amino acid in the encoded protein sequence were removed from the analysis. The secretome prediction was done using SingalP5 (v5.0) (80). Pfam and Gene Ontology (GO) function domains were predicted using InterProScan (v5.42-78.0) (81). Clusters of Orthologous Group (COG) categories were determined for protein sequences using eggNOG-mapper (v2.0) with the taxonomic scope set on Ascomycota (82, 83). Carbohydrate-Active enzymes (CAZymes) were annotated using the dbCAN2 meta server (84, 85). A protein was considered a CAZyme if at least two of the three tools (HMMER,DIAMOND and Hotpep) predicted a CAZyme function.

Parental origin determination

Sub-genomes were divided based on the differences in sequence identities between species A1 and D1/D3 with *V. dahliae*. *V. longisporum* genomes of VLB2, VL20 and PD589 were aligned to the complete genome assembly of *V. dahliae* JR2 using NUCmer (v 3.1), which is

686 part of the MUMmer package (86). Here, only 1-to-1 alignments longer than 10 kb and with a
687 minimum of 80% identity were retained. Subsequent alignments were concatenated if they
688 aligned to the same contig with the same orientation and order as the reference genome. The
689 average nucleotide identity was determined for every concatenated alignment and used to
690 divide the genomes into sub-genomes. Differences in GC-content between homologous genes
691 present in two copies were calculated as described (28). GC content of gene coding regions
692 were calculated with infoseq from EMBOSS (v.6.6.0.0) (87). The features to indicate the
693 biparental origin of the *V. longisporum* genomes were visualized using the R package circlize
694 (v.0.4.10) (88).
695
696 **Genome analysis**
697 The quality of genome assemblies was assessed by screening the presences of Benchmarking
698 Universal Single-Copy Orthologs (BUSCOs) using the BUSCO software version 4.0.6 with
699 the database "ascomycota_odb10" (89).
700 Repeats were *de novo* identified using RepeatModeler (v1.0.11) and combined with
701 the repeat library from RepBase (release 20170127) (90). The genomic location of repeats
702 was identified with RepeatMasker (v4.0.6).
703 The phylogenetic relationship of the nuclear and mitochondrial (sub-)genomes of the
704 *Verticillium* species of the clade Flavnonexudans (38), using following haploid strains: *V.*
705 *alfalfae* = PD683, *V. dahliae* = JR2, *V. nonalfalfae* = TAB2 and *V. nubilum* = PD621 (45,
706 46). Phylogenetic trees based on nuclear DNA were constructed based on the Ascomycete
707 BUSCOs that were shared by all the included species (89). Nucleotide sequences were
708 separately aligned using MAFFT (v7.464) (91). Phylogenetic trees were inferred using
709 RAxML with the GTRGAMMA substitution model (v8.2.11) (92). The robustness of the
710 inferred phylogeny was assessed by 100 rapid bootstrap approximations.

711 Homologs in *Verticillium* were determined using nucleotide BLAST (v2.2.31+).
712 Genes with a minimum identity of 80% and a minimum overlap of 80% were considered
713 homologs, which was determined using the SiLiX (v.1.2.10-p1) software (93).
714 Global nucleotide alignments using the Needle-Wunsch algorithm of the EMBOSS
715 package were used to determine homologous gene pairs in VLB2 and VL20 (v6.6.0.0) (87).
716 Sequence identity between these genes in copy were determined based on their global
717 alignment. Synteny between the VLB2 and VL20 genome assemblies was determined by
718 using one-to-one alignments obtained by NUCmer (v 3.1), which is part of the MUMmer
719 package (86). The synteny was visualized with the R package circlize (v.0.4.10) (88).
720
721 **Gene divergence**
722 Previously published annotations of the haploid *Verticillium* species *V. dahliae*, *V. alfalfae*,
723 *V. nonalfalfae*, *V. nubilum* , *V. tricorpus* and *V. albo-atrum* were used to compare the
724 evolutionary speed of orthologs (45, 46). The VESPA (v1.0b) software was used to automate
725 this process (94). The coding sequences for each *Verticillium* species were filtered and
726 subsequently translated using the VESPA 'clean' and 'translate' function. Homologous genes
727 were retrieved by protein BLAST (v2.2.31+) querying a database consisting of all
728 *Verticillium* protein sequences. Here, the options "-max_hsps 1" and "-qcov_hsp_perc 80"
729 were used. Homologous genes were grouped with the VESPA 'best_reciprocal_group'
730 function. Only homology groups that comprised a single representative for every *Verticillium*
731 spp. were used for further analysis. Protein sequences of each homology group were aligned
732 with muscle (v3.8) (95). The aligned protein sequences of the homology groups were
733 conversed to nucleotide sequence by the VESPA 'map_alignments' function. The alignments
734 were used to calculate *Ka/Ks* for every branch of the species phylogeny using codeml module
735 of PAML (v4.9) with the following parameters: F3X4 codon frequency model, wag.dat

736 empirical amino acid substitution model and no molecular clock (96). To this end, this
737 phylogenetic tree topology was used: ((((*V. dahliae*/D1/D3,(*V. alfalfae*, *V.*
738 *nonalfalfae*)),A1),*V. nubilum*),(*V. tricorpus*, *V. albo-atrum*)). Divergence was only compared
739 for genes that are present in the two sub-genomes of the *V. longisporum* strains VLB2, VL20
740 and PD589.
741
742 **Gene expression analysis**
743 The RNA sequencing reads were filtered using the Trinity software (v2.9.1) option
744 trimmomatic under the standard settings (97). The reads were then mapped to the *Verticillium*
745 genomes using Bowtie 2 (v2.3.5.1) with the first 15 nucleotides on the 5'-end of the reads
746 being trimmed because of inferior quality (98). To compare gene expression patterns,
747 homologs were retrieved by nucleotide blast BLAST (v2.2.31+). Genes with a minimum
748 identity of 80% and a minimum overlap of 80% were considered homologs, which was
749 determined using the SiLiX (v.1.2.10-p1) software (93). Reads were counted to the predicted
750 gene coding regions using the R package Rsubread (v1.34.7) Significant differential
751 expression of a locus was calculated using the R package edgeR (v3.26.8) (99). Significance
752 of differential expression was calculated using t-tests relative to a threshold of log2 fold
753 change of 1 with Benjamini-Hochberg correction using a *p*-value cut-off of 0.05.
754
755 **Data accession**
756 Raw RNAseq reads and genome assemblies are deposited at NCBI under the BioProject
757 PRJNA473305.

Acknowledgements

The authors would like to thank the Marie Curie Actions program of the European Commission that financially supported the research of J.R.L.D. Work in the laboratories of B.P.H.J.T. and M.F.S is supported by the Research Council Earth and Life Sciences (ALW) of the Netherlands Organization of Scientific Research (NWO). B.P.H.J.T acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2048/1 – Project ID: 390686111. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Sander Y.A. Rodenburg for sharing bioinformatics scripts.

References

1. Barton NH. 2001. The role of hybridization in evolution. Mol Ecol 10:551–568.
2. Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities. Annu Rev Genet 45:331–355.
3. Blanckaert A, Bank C. 2018. In search of the Goldilocks zone for hybrid speciation. PLoS Genet 14:e1007613.
4. Mallet J. 2007. Hybrid speciation. Nature 446:279–283.
5. Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat Ecol Evol 3:170-177.
6. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MHY, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušic I, Doronichev VB, Golovanova L V., Lalueza-Fox C, De La Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.
7. Stukenbrock EH. 2016. The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology 106:104–112.
8. Depotter JRL, Seidl MF, Wood TA, Thomma BPHJ. 2016. Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. Curr Opin Microbiol 32:7–13.
9. Gabaldón T. 2020. Hybridization and the origin of new yeast lineages. FEMS Yeast Res 20:foaa040.
10. Leducq AJ, Nielly-thibault L, Charron G, Verta J, Samani P, Sylvester K, Hittinger CT, Bell G, Landry CR. 2015. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. Nat Microbiol 1:15003.
11. Marcet-Houben M, Gabaldón T. 2015. Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. PLOS Biol 13:e1002220.
12. Mixão V, Gabaldón T. 2020. Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*. BMC Biol 18:48.
13. Mixão V, Hansen AP, Saus E, Boekhout T, Lass-Florl C, Gabaldón T. 2019. Whole-genome sequencing

of the opportunistic yeast pathogen *Candida inconspicua* uncovers its hybrid origin. Front Genet 10:383.

14. Pryszcz LP, Németh T, Saus E, Ksiezopolska E, Hegedűsová E, Nosek J, Wolfe KH, Gacser A, Gabaldón T. 2015. The genomic aftermath of hybridization in the opportunistic pathogen *Candida metapsilosis*. PLoS Genet 11:e1005626.

15. Schröder MS, Martinez de San Vicente K, Prandini THR, Hammel S, Higgins DG, Bagagli E, Wolfe KH, Butler G. 2016. Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. PLoS Genet 12:e1006404.

16. Cook DE, Mesarich CH, Thomma BPHJ. 2015. Understanding plant immunity as a surveillance system to detect invasion. Annu Rev Phytopathol 53:541–563.

17. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, Zuccaro A, Reissmann S, Kahmann R. 2015. Fungal effectors and plant susceptibility. Annu Rev Plant Biol 66:513–545.

18. Rovenich H, Boshoven JC, Thomma BPHJ. 2014. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. Curr Opin Plant Biol 20:96–103.

19. Snelders NC, Kettles GJ, Rudd JJ, Thomma BPHJ. 2018. Plant pathogen effector proteins as manipulators of host microbiomes? Mol Plant Pathol 19:257–259.

20. Snelders NC, Rovenich H, Petti GC, Rocafort M, van den Berg GCM, Vorholt JA, Mesters JR, Seidl MF, Nijland R, Thomma BPHJ. 2020. Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. Nat Plants 6:1365–1374.

21. Li F, Upadhyaya NM, Sperschneider J, Matny O, Nguyen-Phuc H, Mago R, Raley C, Miller ME, Silverstein KAT, Henningsen E, Hirsch CD, Visser B, Pretorius ZA, Steffenson BJ, Schwessinger B, Dodds PN, Figueroa M. 2019. Emergence of the Ug99 lineage of the wheat stem rust pathogen through somatic hybridisation. Nat Commun 10:5068.

22. Singh RP, Hodson DP, Jin Y, Lagudah ES, Ayliffe MA, Bhavani S, Rouse MN, Pretorius ZA, Szabo LJ, Huerta-Espino J, Basnet BR, Lan C, Hovmøller MS. 2015. Emergence and spread of new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control. Phytopathology 105:872-884.

23. Menardo F, Praz C, Wyder S, Bourras S. A, McNally KE, Parlange F, Riba A, Roffler S, Schaefer L, Shimizu KK, Valenti L, Zbinden H, Wicker T, Keller B. 2016. Hybridization of powdery mildew strains gives raise to pathogens on novel agricultural crop species. Nat Genet 48:201–205.

24. Runemark A, Vallejo-Marin M, Meier JI. 2019. Eukaryote hybrid genomes. PLoS Genet 15:e1008404.

827 25. McClintock B. 1984. The significance of responses of the genome to challenge. Science 226:792–801.
828 26. Matute DR, Butler IA, Turissini DA, Coyne JA. 2010. A test of the snowball theory for the rate of
829 evolution of hybrid incompatibilities. Science 329:1518–1521.
830 27. Mixão V, Gabaldón T. 2017. Hybridization and emergence of virulence in opportunistic human yeast
831 pathogens. Yeast 35:5–20.
832 28. Louis VL, Despons L, Friedrich A, Martin T, Durrens P, Casarégola S, Neuvéglise C, Fairhead C, Marck
833 C, Cruz JA, Straub M-L, Kugler V, Sacerdot C, Uzunov Z, Thierry A, Weiss S, Bleykasten C, De
834 Montigny J, Jacques N, Jung P, Lemaire M, Mallet S, Morel G, Richard G-F, Sarkar A, Savel G,
835 Schacherer J, Seret M-L, Talla E, Samson G, Jubin C, Poulain J, Vacherie B, Barbe V, Pelletier E,
836 Sherman DJ, Westhof E, Weissenbach J, Baret P V., Wincker P, Gaillardin C, Dujon B, Souciet J-L.
837 2012. *Pichia sorbitophila*, an interspecies yeast hybrid, reveals early steps of genome resolution after
838 polyploidization. G3 2:299–311.
839 29. Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH. 2012. Fusion of two divergent
840 fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc Natl Acad
841 Sci 109:10954–10959.
842 30. Ortiz-Merino RA, Kuanyshev N, Braun-Galleani S, Byrne KP, Porro D, Branduardi P, Wolfe KH. 2017.
843 Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a
844 failed mating-type switch. PLOS Biol 15:e2002128.
845 31. Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nat Rev
846 Genet 18:411-424.
847 32. Hellsten U, Khokha MK, Grammer TC, Harland RM, Richardson P, Rokhsar DS. 2007. Accelerated
848 gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. BMC Biol 5:31.
849 33. Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression
850 bias and expression level dominance in allopolyploids. New Phytol 196:966–971.
851 34. Zhang M, Tang YW, Qi J, Liu XK, Yan DF, Zhong NS, Tao NQ, Gao JY, Wang YG, Song ZP, Yang J,
852 Zhang WJ. 2019. Effects of parental genetic divergence on gene expression patterns in interspecific
853 hybrids of *Camellia*. BMC Genomics 20:828.
854 35. Yoo M-J, Szadkowski E, Wendel JF. 2013. Homoeolog expression bias and expression level dominance
855 in allopolyploid cotton. Heredity 110:171–180.
856 36. Cox MP, Dong T, Shen G, Dalvi Y, Scott DB, Ganley ARD. 2014. An interspecific fungal hybrid

reveals cross-kingdom rules for allopolyploid gene expression patterns. PLoS Genet 10:e1004180.

37. Praz CR, Menardo F, Robinson MD, Müller MC, Wicker T, Bourras S, Keller B. 2018. Non-parent of origin expression of numerous effector genes indicates a role of gene regulation in host adaption of the hybrid triticale powdery mildew pathogen. Front Plant Sci 9:49.

38. Inderbitzin P, Bostock RM, Davis RM, Usami T, Platt HW, Subbarao K V. 2011. Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. PLoS One 6:e28341.

39. Inderbitzin P, Davis RM, Bostock RM, Subbarao K V. 2011. The ascomycete *Verticillium longisporum* is a hybrid and a plant pathogen with an expanded host range. PLoS One 6:e18260.

40. Depotter JRL, Seidl MF, van den Berg GCM, Thomma BPHJ, Wood TA. 2017. A distinct and genetically diverse lineage of the hybrid fungal pathogen *Verticillium longisporum* population causes stem striping in British oilseed rape. Environ Microbiol 19:3997-4009.

41. Depotter JR, Deketelaere S, Inderbitzin P, Tiedemann AV, Höfte M, Subbarao KV, Wood TA, Thomma BPHJ. 2016. *Verticillium longisporum*, the invisible threat to oilseed rape and other brassicaceous plant hosts. Mol Plant Pathol 17:1004-1016.

42. Inderbitzin P, Subbarao K V. 2014. *Verticillium* systematics and evolution: how confusion impedes Verticillium wilt management and how to resolve it. Phytopathology 104:564–574.

43. Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert JP, Noble WS. 2015. Accurate identification of centromere locations in yeast genomes using Hi-C. Nucleic Acids Res 43:5331-5339.

44. Seidl MF, Kramer HM, Cook DE, Fiorin GL, van den Berg GCM, Faino L, Thomma BPHJ. 2020. Repetitive elements contribute to the diversity and evolution of centromeres in the fungal genus *Verticillium*. MBio 11:e01714-20.

45. Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF. 2018. Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangements and gene losses. Environ Microbiol 20:1362–1373.

46. Faino L, Seidl M, Datema E, van den Berg GCM, Janssen A, Wittenberg AHJ, Thomma BPHJ. 2015. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome. MBio 6:e00936-15.

47. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. Methods Mol Biol 1962:65-95.

887 48. Seidl MF, Thomma BPHJ. 2014. Sex or no sex: evolutionary adaptation occurs regardless. BioEssays
888 36:335–345.
889 49. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma BPHJ.
890 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal
891 pathogen. Genome Res 26:1091–1100.
892 50. Kimura M, Ohta T. 1974. On some principles governing molecular evolution. Proc Natl Acad Sci
893 71:2848–2852.
894 51. Depotter JRL, Rodriguez-Moreno L, Thomma BPHJ, Wood TA. 2017. The emerging British
895 *Verticillium longisporum* population consists of aggressive *Brassica* pathogens. Phytopathology
896 107:1399-1405.
897 52. Lanver D, Tollot M, Schweizer G, Lo Presti L, Reissmann S, Ma LS, Schuster M, Tanaka S, Liang L,
898 Ludwig N, Kahmann R. 2017. *Ustilago maydis* effectors and their impact on virulence. Nat Rev
899 Microbiol 15:409–421.
900 53. Gagic D, Ciric M, Wen WX, Ng F, Rakonjac J. 2016. Exploring the secretomes of microbes and
901 microbial communities using filamentous phage display. Front Microbiol 7:429.
902 54. Li XC, Fay JC. 2017. Cis-regulatory divergence in gene expression between two thermally divergent
903 yeast species. Genome Biol Evol 9:1120-1129.
904 55. Hovhannisyan H, Saus E, Ksiezopolska E, Gabaldón T. 2020. The transcriptional aftermath in two
905 independently formed ybrids of the opportunistic pathogen *Candida orthopsilosis* . mSphere 5:e00282-
906 20.
907 56. Hovhannisyan H, Saus E, Ksiezopolska E, Hinks Roberts AJ, Louis EJ, Gabaldón T. 2020. Integrative
908 omics analysis reveals a limited transcriptional shock after yeast interspecies hybridization. Front Genet
909 11:404.
910 57. Shi X, Ng DWK, Zhang C, Comai L, Ye W, Chen ZJ. 2012. Cis- and trans-regulatory divergence
911 between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nat
912 Commun 3:950.
913 58. Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of
914 gene expression regulation. Science 324:659–662.
915 59. de Jonge R, Bolton MD, Kombrink A, Van Den Berg GCM, Yadeta KA, Thomma BPHJ. 2013.
916 Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. Genome Res

23:1271–1282.

60. Cook DE, Kramer HM, Torres DE, Seidl MF, Thomma BPHJ. 2020. A unique chromatin profile defines adaptive genomic regions in a fungal plant pathogen. Elife 9:e62208.

61. Depotter JRL, Shi-Kunne X, Missonnier H, Liu T, Faino L, van den Berg GCM, Wood TA, Zhang B, Jacques A, Seidl MF, Thomma BPHJ. 2019. Dynamic virulence-related regions of the plant pathogenic fungus *Verticillium dahliae* display enhanced sequence conservation. Mol Ecol 28:3482-3495.

62. Klosterman SJ, Subbarao K V, Kang S, Veronese P, Gold SE, Thomma BPHJ, Chen Z, Henrissat B, Lee Y-H, Park J, Garcia-Pedrajas MD, Barbara DJ, Anchieta A, de Jonge R, Santhanam P, Maruthachalam K, Atallah Z, Amyotte SG, Paz Z, Inderbitzin P, Hayes RJ, Heiman DI, Young S, Zeng Q, Engels R, Galagan J, Cuomo C a, Dobinson KF, Ma L-J. 2011. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog 7:e1002137.

63. de Jonge R, van Esse HP, Maruthachalam K, Bolton MD, Santhanam P. 2012. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc Natl Acad Sci 109:5110–5115.

64. Kombrink A, Rovenich H, Shi-Kunne X, Rojas-Padilla E, van den Berg GCM, Domazakis E, de Jonge R, Valkenburg DJ, Sánchez-Vallet A, Seidl MF, Thomma BPHJ. 2017. *Verticillium dahliae* LysM effectors differentially contribute to virulence on plant hosts. Mol Plant Pathol 18:596–608.

65. Lynch M, Conery JS. 2001. The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.

66. Sriswasdi S, Takashima M, Manabe R, Ohkuma M, Sugita T, Iwasaki W. 2016. Global deceleration of gene evolution following recent genome hybridizations in fungi. Genome Res 26:1081–1090.

67. Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr Opin Plant Biol 15:147–153.

68. Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. Proc Natl Acad Sci 102:5454–5459.

69. Smukowski Heil CS, DeSevo CG, Pai DA, Tucker CM, Hoang ML, Dunham MJ. 2017. Loss of heterozygosity drives adaptation in hybrid yeast. Mol Biol Evol 34:1596-1612.

70. Hou J, Friedrich A, De Montigny J, Schacherer J. 2014. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. Curr Biol 24:1153-1159.

71. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu□: scalable and

accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736.

72. Firtina C, Kim JS, Alser M, Cali DS, Cicek AE, Alkan C, Mutlu O. 2020. Apollo□: a sequencing-technology-independent, scalable, and accurate assembly polishing algorithm. Bioinformatics 36:3669-3679.

73. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997.

74. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3:95-98.

75. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356: 92-95.

76. Dudchenko O, Shamim M, Batra S, Durand N, Musial N, Mostofa R, Pham M, Glenn St Hilaire B, Yao W, Stamenova E, Hoeger M, Nyquist S, Korchina V, Pletch K, Flanagan J, Tomaszewicz A, McAloose D, Pérez Estrada C, Novak B, Omer A, Aiden E. 2018. The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000. bioRxiv:10.1101/254797.

77. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

78. Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100.

79. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2□: accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions. Genome Biol 14:R36.

80. Juan J, Armenteros A, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Heijne G Von, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423.

81. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.

82. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol

34:2115-2122.

83. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P. 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309-D314.

84. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. DbCAN2: A meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95-W101.

85. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. DbCAN: A web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445-W451.

86. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12.

87. Rice P, Longden I, Bleasby A. 2000. EMBOSS□: the European molecular biology open software suite. Trends Genet 16:276–277.

88. Gu Z, Gu L, Eils R, Schlesner M, Brors B. 2014. Circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812.

89. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing genome assembly and annotation completeness. Methods Mol Biol 1962:227-245.

90. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11.

91. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.

92. Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

93. Miele V, Penel S, Duret L. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics 12:116.

94. Webb AE, Walsh TA, O'Connell MJ. 2017. VESPA: Very large-scale Evolutionary and Selective Pressure Analyses. PeerJ Comput Sci 3:e118.

95. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

96. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591.

97. Grabherr MG., Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W. N, Friedman and AR. 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644–652.

98. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.

99. Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.